Jack Stenner
University of Florida, USA

Gregory L. Ulmer
University of Florida, USA

Mood Themes the World

Abstract: Apparatus theory (a hybrid of McLuhan and Derrida) hypothesizes that a civilization of electracy (the digital apparatus) must learn how to thrive in a lifeworld in which the visceral faculty of appetite is hegemonic. The dominant axis of behavior today is fantasy-anxiety (attraction/repulsion). We propose that world theming has created a vernacular discourse that may be raised to a second power of expression as vehicle of visceral intelligence. The immediate claim is that theming in digital media augments mood (ambiance) into a power of imagination, just as dialectic in writing augmented logic into a power of reason. Fantasy today is persuasive, just as logical entailment is (was) in the rational order of literacy. Decisions determining real events today are being made in worlds of mood.

World theming is evident in the vernacular art practices arising from recent advances in artificial intelligence. The availability of commodity GPUs, along with public access to advanced research via GitHub, Kaggle, Hugging Face, and the proliferation of forums such as Reddit, Discord, YouTube, and others, has resulted in a renaissance of public engagement with technology-informed creative practice. In addition, the general availability of Google's previously internal-only development tool, Colab, in late 2017 provided access to cloud-based GPUs and storage systems accessible only to data scientists and academics.
In early 2021 Ryan Murdock released a Colab notebook called Big Sleep that combined OpenAI's recently published Contrastive Language-Image Pre-training (CLIP) with BigGAN. This model is a paradigmatic example of our observation. By early 2022, multiple derivations of this process incorporated alternative image generation techniques. This paper will demonstrate how the fundamental basis of these methods are distinctly electrate in their use of 'theme' and emphasis on 'mood' in world-building, including a case-study animation called *Dissipative Off-ramps*.

Keywords: apparatus; theme; mood; artificial intelligence; electracy.

Electracy

What is the metaverse for? Electracy as apparatus offers a comprehensive hypothesis. The point of departure is the most familiar use of "apparatus", Louis Althusser's Ideological State Apparatus (ISA). Ideology is defined as the imaginary relation

*Author contact information: stenner@ufl.edu
**Author contact information: glue@ufl.edu

of people to their real conditions of existence. Four institutions are responsible for reproducing education in society, Althusser says: home, church, school, and media. Alain Badiou, Althusser's student, revised this set as love (psychoanalysis), politics (law), science, and art, each institution having its own truth procedure, creating conditions making philosophy possible. The insight of apparatus is that these institutions are part of a 'stack', or 'cognitive map', of corresponding levels from individual human body (the body with its three faculties – head, heart, viscera according to Plato, or theoria, praxis, poiesis according to Aristotle, or Real, Symbolic, Imaginary according to Lacan), through the ISAs to the historical epochs: Paleo, Orality, Literacy, Electracy. 'Theopraxesis' is a portmanteau term naming a practice emerging in digital media of 'thinking-doing-making' with the stack. The argument is that each human capability is individuated transductively (individual and collective emerging together), evolving in sequence, each hegemonic in its own historical epoch, producing a metaphysics (reality) native to that capability. The shift from one phase to the next is a disruptive jump, requiring 'innervation', the invention of metaphysical practices adapting people to the new conditions of their world, which include neurophysiological mutations in the human sensorium itself. Walter Benjamin got in trouble with his Marxist friends for endorsing the early animations of Mickey Mouse, based on their innervating effect, adjusting viewers to the shock of life in industrial cities. The purpose of metaverse is innervation for electracy.

All four apparatus coexist today, each one with a native technology, institution, and identity behavior. The prototype is literacy: oral society acquiring alphabetic writing; Plato opening the first school (the Academy) undertaking the invention of method, the logic augmenting the faculty of reason, and creating the dialogue form dramatizing the performance of selfhood as identity, personified in Socrates. Plato theorized his world production in *Timaeus*: a chora (space of generation) sorts out the chaos of material becoming into the formal being of the four elements (Earth, Air, Fire, Water). Each apparatus constitutes a chora, a civilizational interface sorting chaos into cosmos (chaosmos). Oral Greece assumed divine causality, managed by right-wrong behavior, religion interfacing with the will of gods by means of ritual and mythology. Literacy was a new metaphysics whose causal principle was nature (*Phusis*), speaking mathematics, managed by true-false propositional logic, science opening a new dimension of reality. The execution of Socrates for the crime of impiety stands as a monument to the antagonism between these apparatus faculties, marked again in the trial of Galileo, continuing today in such events as the 9/11 attacks.

Science producing ultimately the technology of the industrial revolution proved to be a pharmakon, as Bermard Stiegler warned, dephasing into electracy (digital apparatus) with the poison/gift of the Anthropocene, beginning in the late eighteenth century. The Anthropocene, in which human activity begins to influence the planetary environment, identifies the causal principle of electracy as the human faculties themselves. Electracy is the apparatus of visceral appetite (sensation, poiesis). The institution native to electracy is the corporation, inscribing capitalism with the

supporting mathematical inventions of probability and the stock market to manage risk (fruits of ambitious gamblers). The creation of the commodity form, the separation of exchange value from use value, supported the new behaviour of consumption, as advertising exploited media inventions to access human libido, locus of the causal power, the force or energy of electracy. Advertisers learned that what matters now is not the steak (substance) but the sizzle (style).

The academy of electracy opened in Montmartre in Paris, the cabaret vice district, beginning in the 1880s, culminating in Zurich's Cabaret Voltaire during World War I with the invention of Dadaism. The vanguard arts revolution created the operating logics needed to innervate the chaos of the industrial city – collage montage, assemblage. Paradigm of the operating devices of electracy is Cézanne's 'little sensation' of passage – creating dimensionality with the perceptual effect of the human eye, that warm colors advance, and cool colors recede. Proust's involuntary memory is another example, or the poetics of symbolist epiphany in general (Mallarmé's flower absent from all bouquets), doing for perceptual functioning of felt sensation what the noncontradiction principle of written logic did for reason. Dadaism was created as entertainment, in the same venues experimenting with the entertainment possibilities of media technologies. Entertainment as a syncretism of high and pop arts spread through mediated cities, leading to themed environments, with theming being to the dissemination of electracy within literacy what secularization was to the dissemination of literacy within oral culture. Through Cézanne artists realized art plays not with the look of nature but with the human sensorium. In electracy, being is made present through the formal properties of style.

Las Vegas is an heir of the cabaret scene, the casinos themed as a fantasized city (Venice in the Venetien) being like Gothic cathedrals of the visceral apparatus, performing the rituals of capitalist gamblers. The prototype of theming is Disneyland, the *Republic* of electracy, with Walt Disney earning designation as our Plato by his mastery of the 'plasmatic line' in his animated cartoons, translating the formal device of little sensations into mass entertainment. Disney helps us understand that the new category sorting chaos into order created in the cabaret is the *simulacrum*, defined as a copy without original. It functions as caricature, selecting just a few traits to produce a hyperreal effect: *more real than reality itself*. Elvis impersonation is a paradigm of caricature: a singer, themed as Elvis, needs only three traits to be universally recognizable: lip curl, hip swivel, leg bounce. The device correlates with how human memory works, and as operated in digital media augments the faculty of imagination. As Kant explained in his third critique on the Judgment of Taste, the experience of the sublime (simultaneous attraction and repulsion in the human encounter with overwhelming hyperobjects) is due to disparity between the power of reason (augmented in literacy) and the impotence of imagination (loss of intuition) in need of its own augmentation. Electracy is doing for attraction-repulsion visceral emotion, what literacy did for true-false logic, or orality for right-wrong behavior, or paleo for life-death survival (the four coexisting metaphysical realities).

Electracy augments imagination, to bring intuition back into balance with the faculties of reason and will–precisely the function Kant proposed for aesthetics, bridging the abyss separating science and religion. Hollywood mediatized the themed environment, taking over from Architecture the function of monumentality (collective identity formation), reorienting dwelling to manage general sprawl blurring all categories of literacy. Paul Virilio named this condition *dromosphere*, the dimension collapse resulting from real time light-speed technology. Disneyland models utopian urban planning in the sprawl of southern California, just as it organizes the categorial sprawl of imagination through theming, evolving into the global capital of ecstacity. The chora of electracy instantiated in media production merged direct and represented experience of environment, leading to what the architect Nigel Coates called 'ecstacity'. Transmedia franchise brands are industrializing ecstacity, blending real and fictional worlds. Ideology itself (as the imagined experience of real conditions) is being revised into the diegeses of fictional worlds, creating new identity behaviours (epitomized in cosplay) native to electracy. For tourists, Oxford becomes the world of Harry Potter, or New Zealand the world of Hobbits.

The elements for producing a cognitive map (Existential Positioning System) orienting netizens to the stack of theopraxesis are available in vernacular electracy emerging in transmedia worlds. One of the telling facts of our epoch is that cinema and psychoanalysis appeared in the same year (1895), each innervating a dimension of visceral capability (cinematic imaging articulates unconscious desire, just as written syllogisms articulated rational cognition or performed rituals articulated behavioural belief). Cinema was invented through learning how to tell a story in film. The process developed multiple layers of coherence, from story at the highest level of archetypal familiarity, adding further layers of coherence with narrative form, character performance, image design, and musical pattern, gathering now in world mood. It is perhaps counterintuitive to say that world is the *least* organized of the series, evoked as atmosphere, a feeling, an intuition of how we want life to be in the modality of impossible (fantasy). Media augmentation of the formal devices of vanguard aesthetics, the little sensations, trigger unconscious involuntary experience in which fantasy is defense against anxiety (attraction-repulsion). Described as the architecture of reassurance, Disneyland functioned as antidote to Cold War anxiety.

Our hypothesis is that world theming has created a vernacular discourse that may be raised to a second power of theopraxesis, in service of innervation, as philosophy raised written epics into truth, or poetry raises everyday language into beauty. The educational project does not stop with adaptation to a visceral consumer society but adds the imperative to 'temper' the choras, attune the four realities into a harmonious 'music of the spheres'. The immediate claim is that theming in digital media augments mood (ambiance) into a power of imagination, just as dialectic in writing augmented logic into a power of reason. Fantasy touching sensation is persuasive, just as logical entailment is persuasive in the rational order. Decisions determining real events today are being made in worlds of mood. Metaverse is emerging as electrate chora, a site of

innervation, whose purpose is adapting the human sensorium to real time conditions of dromosphere. In metaverse, humanity engages with itself as world demiurge.

Involuntary ontology

The graphic novelist and educator Lynda Barry orients us to the pedagogy of our digital experiment. Her fundamental criterion for drawing is 'aliveness', which is to say she shows how drawing may function as ontology. In *Syllabus,* her graphic novel of curriculum and pedagogy, she describes the exercise in which students were asked to draw a car, and then Batman. She observed that the sketches by those who hadn't drawn since childhood had a quality of aliveness, unlike the ones by those relying on conventional training. This formal expressive quality of aliveness – the plasmatic depictive line – has been valued since Chauvet cave. Hogarth's S curve of beauty is its monogram, and the sine wave marks its grounding in physics. Exemplary of theopraxesis, Barry's exercise for working with memory is to divide a page into four sections, imagine it as framing a scene past, present, or future. In this frame annotate what is heard (verbal), seen (visual), performed (enacted), plus a drawing (sketch). We recognize in this inventory the elements of Joyce's epiphanic scenes, which he defined in *Stephen Hero* as "a sudden spiritual manifestation, whether in the vulgarity of speech or of gesture or of a memorable phase of the mind itself." Her aliveness correlates with Joyce's adaptation of Aquinas's three phases of artistic apprehension: *integritas, consonantia, claritas*. Barry's annotations and Joyce's aesthetics exemplify correlated human capabilities, assumed in all manner of interface from Donald Norman's experience design (useful, usable, desirable) to the Graphical User Interface (keyboard, mouse, windows). These criteria of aesthetic aliveness guide design of a figure (proportional analogy) bringing into relation makers' embodiment with documents of the cultural archive, articulating an event of attraction-repulsion. The power of art to correlate faculties for an individual, opens the question of whether distributed annotations in metaverse might catalyse collective epiphany.

Barry's exercise resonates with the philosophers asking, *what makes life worth living*? This is the question driving our experiment, to undergo the experience of aliveness. We recall the paradigmatic scene of Proust's involuntary memory, biting into the madeleine, the tea biscuit, catalyzing a memory of his beloved grandmother, experiencing a happiness so profound that he spent the rest of his life searching for its source, as recorded in his novel, *In Search of Lost Time*. This memory is dramatized in the novel by the effect on Charles Swann, captivated by the 'little phrase' in the sonata by Vinteuil, whose aesthetic power activates a capability of feeling enabling Swann to fall in love when he meets Odette. The first chapter of *Flow: The Psychology of Optimal Experience*, by Mihaly Csikszentmihalyi, is entitled "Happiness Revisited". *Flow* makes explicit the role of mood in creativity. The lesson is that world-building happens in the middle voice, autotelic or autotropic, self-affecting, driven by the feeling of optimal experience, regardless of the semantics of the problem being solved or invention discovered, or work designed. The name for the feeling of being alive is

happiness against anxiety. As we know, advertisers are the sophists of happiness, but to their credit they have mastered the craft of the little sensation catalyzing fantasy, thus creating a vernacular electracy.

While the corporations themselves seem uncertain about a commercial formula for 'world sustainability', a vernacular version already exists in the aesthetics of traditional folk cultures, teaching that the fundamental category of world has always been mood. Roland Barthes knew what he was doing when he devoted an entire semester of his final year-long seminar to haiku poetry. He had already argued in *Camera Lucida* that haiku offers a poetics of photography, the snapshot as a machine for recording immediate moments. Barthes wanted to understand how the aphorism of thought and the anecdote of life converged – how to 'write' the world as he was living it (his pleasure of the text). The second semester was devoted to Proust. Haiku helped Barthes understand what he called the third or obtuse meaning, the punctum or sting of involuntary memory received from some random detail in an image. Aliveness is this unforeseeable spark between a formal material detail and a personal memory. Punctum is the fundamental operator of electrate metaphysics. Metaverse needs to augment punctum at scale.

Considering the centrality of aesthetics to electracy, Japanese arts are a major resource. The traditional cultural mood of wabi-sabi (the rustic), for example, demonstrates how a simple image may become wide (sustain a world). The iconic instance is the mood associated with cherry blossoms, when their peak flowering passes and they begin to fall and become litter. An old iron nail making a rusty streak in a weathered shingle expresses an entire worldview, for those native to Buddhist wisdom. 'Aliveness' is this capacity of a material detail to open a circuit of extimacy between environment and perception. The fallen blossoms are the vehicle, and the themata or tenor is wisdom, formulated as answers to the three questions of theopraxesis, catechism of human capabilities: *Metaphysics*: what is the nature of the world? *Morality*: given this nature, how should I act? *Mood*: given this action in such a world, how do I feel (Truth, Goodness, Beauty). Each wisdom tradition has its own answers, just as each person intuitively does.

Wabi-Sabi as a melancholic awareness of temporality has its equivalent in other folk cultures, developed into sophisticated creoles in the black Atlantic: blues/jazz, saudade/samba, mufarse/tango. Russian *nostalghia* or Spanish *duende* are old world versions, resonant with *han* in Korea. The feeling has been summarized as *glad to be feeling sad* (a bittersweet taste). A paradigmatic instance of transduction, of simultaneous individuation of an individual and a city, is Orhan Pamuk's *Istanbul*. The Nobel Laureate's account of the melancholy of his *associated milieu* amounts to a document of Turkish wabi-sabi, known as *hüzün*. Anne Carson in her book, *Eros the Bittersweet*, helps identify this feeling as love itself. If we understand folk culture as a certain preservation of collective optimal experience, it clarifies why mood, rather than reason or belief, is the defining quality of world, functioning as ontological – life (being-in-the-world) made present.

How is aliveness supported in metaverse? How is the axis of attraction-repulsion materialized as ethics and politics? Can metaverse accept the challenge of innervating the dromosphere by means of entertainment? The challenge is a paradigm shift, to reframe the Anthropocene within a new metaphysics. Our ethics and politics have shifted to the chora of fiction and fantasy. Plato included in the *Republic* a model of education appropriate for an ideal literate city: the allegory of the cave. A found allegory of electrate education is the comic strip, *Krazy Kat*, by George Herriman, running in Hearst newspapers from 1913 to 1944. The variations played on the relationship among Ignatz Mouse, Krazy Kat, Officer Pupp, and the brick, diagram the dynamics of a *universal motor* driving digital ontogenesis. The four actants of this drama model the primordial forces sorted out from chaos, and are found in every dimension of stack, as for example the four suits of cards in Tarot's Minor Arcana, or the elements of every game played with a ball, or even the four forces of physics or four tastes (with umami as the sphinx).

In his semiophysics René Thom, creator of catastrophe topology, used cat-mouse predation as model of life dynamics. World is a field which both predator and prey fill with *pregnance* (affective emotion of hunger, fear, sexual need). This affective *pregnance* is primary in electracy, equivalent to conceptual substance in literacy. The ground of this field consists of noise (chaos), characterized by Serres as 'parasite', which transforms into signal when the intentional object introduces a disturbance, creating a new pattern (salience) – figure/ground aspect shift being a basic feature of human perception. Serres' fable of modern information cybernetics is the story of the Lamb, Shepherd, and Wolf, precisely the kind of story circulating as wisdom from Aesop to Disney.

Thom diagrams the chase scene as the cusp catastrophe, a phase shift generated by the hunt, the subject conjoining or disjoining with the object of desire. The Subject's desire for the Object (whatever it may be – a miser's love of gold; Pygmalion's passion for Galatea, Geppetto and Pinocchio, Proust's Swann and Odette) is ontological, the source of aliveness in electracy. Here is electrate cause: *passion*. As every cat-mouse chase cartoon teaches, the prey (slave) is as cunning as the predator (master). Matter and form are both actants: enjoyment is the chase itself. Jacques Lacan's notion of the gaze informing the field theory of subjectivation, is based on Roger Caillois's account of mimicry across all life, the play between caricature (exaggeration) and camouflage (deception), in the drama of survival between predator and prey (*conatus*, the will to live as the first apparatus). Gilles Deleuze and Felix Guattari's rhizome, their figure for relational ontology, takes as its prototype the mimetic play between the wasp and the orchid, in which the orchid tricks the wasp into the role of sex partner. Lacan theorized human libido, the energy of electrate metaphysics, as an asymmetrical tension between aim and goal, whose necessary failure sustains life. This drama is played out on the US/Mexico border, with Migrant, Coyote, ICE, and Wall. An immediate takeaway concerns the world opening in metaverse (scene of the gaze): it is predatory play. AI metaphysics is a chase cartoon.

Theme

World theming is evident in the vernacular art practices arising from recent advances in artificial intelligence. The availability of commodity GPUs, along with public access to advanced research via GitHub, Kaggle, Hugging Face, and the proliferation of forums such as Reddit, Discord, YouTube, and others, has resulted in a renaissance of public engagement with technology-informed creative practice. For example, Google's DeepDream[1] inspired generative artists to explore the latent space of convolutional neural networks to creative, hallucinatory ends. In addition, the public availability of Google's previously internal-only development tool, Colab, in late 2017 provided access to cloud-based GPUs and storage systems accessible only to data scientists and academics.

For this project, we tested the principles of Cézanne's 'little sensation' or Proust's 'little phrase' as poetics of motif, proposing that 'motif' is to electracy what 'concept' is to literacy. In music, a motif is "a short succession of notes producing a single impression; a brief melodic or rhythmic formula out of which longer passages are developed." In biochemistry, a motif is, "a distinctive sequence on a protein or DNA" (Oxford). In essence, a motif is a recognizable pattern. These patterns can be assembled and manipulated to produce a composition or a living organism. This theme can become associated with an artist as their 'style'. Other artists can apply this theme to their work in varying degrees by repurposing the motifs used in the original. A classic example is the Picasso-derived Cubist aspects of Duchamp's *Nude Descending a Staircase* (1912).

The key feature of a theme is that we use the sequence of patterns as a device to represent something else, a larger concept or a feeling. The bridges, brickwork, and arches that let us recognize Venice become a theme when a casino is represented AS Venice in Las Vegas. The production and utilization of machine learning models prevalent in artificial intelligence (AI) invoke similar processes of theming that are unique to electracy. Artists traverse the latent space produced through the training process to reveal motifs that emerge from the theme of the model.

Machine learning is built around the concept that we can mathematically identify patterns in datasets and encapsulate these relationships in models that can be used to efficiently retrieve probabilities that new input data is similarly related. To work with AI is to work with a motif identification system. AI machine learning models attempt to define a function that provides the best correlation between a set of inputs and outputs. We choose algorithms that are most likely to produce accurate predictions for the task at hand. For example, a financial model might be trained on historic trading behavior or geographic housing data. A language model might be trained to predict the next word in a poem or the most likely medical term to describe a set of symptoms. In visual art, the primary models are, of course, vision based.

---

[1] Alexnader Mordvintsev, "Inceptionism: Going Deeper into Neural Networks," *Google AI Blog*, June 17, 2015, http://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html, acc. on March 6, 2023.

Vision models

Vision models have existed in artificial intelligence, specifically the subcategory called machine learning, for decades. Convolutional Neural Networks (CNN) were introduced in the late 1980s. By the mid-2010s, advances in GPU technology and new data science techniques resulted in the development of AI vision models that involve billions of parameters and images. These models range from Autoregressive and Generative Adversarial Networks to the latest Transformer and Diffusion based techniques, each competing to be the most comprehensive, fastest, and most accurate technique. Each of these vision models leverages publicly and privately assembled image datasets data scientists use to 'pre-train' AI vision models for use in subsequent work. Researchers pre-train their models with popularly available datasets so scientific and performance comparisons can be made between their work and those working in associated domains. There is debate over the importance of the size of a dataset or the point at which increased size produces diminishing returns. Google, Meta, OpenAI, and others, working with computer scientists in academia, have leveraged their access to data and datacenters to build models that are massive in scale.

Shoshanna Zuboff,[2] Kate Crawford[3] and many others have written about how these platforms commodify human interaction to the point of directing individual behavior and the perils of granting such power to these corporations. This power, combined with the fact that the data often codifies social biases and viewpoints that do not reflect our goals as humanitarians, is well covered in the literature. The film *Coded Bias*[4] is a good starting point for reference. To recognize bias is to acknowledge that the structures of AI are never neutral. Machine models, including our working processes, thematize the subject. Our intention here is not to repeat these critiques. Still, these concerns raise the stakes on the necessity to understand better how the material and ideological apparatus of electracy is constructing our future. Theming is at work across all aspects of AI, but it is particularly illustrated in the architecture of computer vision models. The earliest vision models emphasized the relationship of image features to calculate probabilities that a subject matched a given input. Images were broken down into smaller pixel arrays and mapped into a 'latent space' based on similarity. Autoregressive models predict future values based on past values seen in the training data. The dataset is used to train the model to work with content of the type contained within the data. A simple example is the MNIST dataset.

---

[2] Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (First Trade Paperback Edition, PublicAffairs, 2020).

[3] Kate Crawford, *Atlas of Ai: Power, Politics, and the Planetary Costs of Artificial Intelligence* (Yale: Yale University Press, 2021).

[4] S. Kantayya (Director), *Coded Bias*, 2020, https://www.codedbias.com, acc. on March 6, 2023.

Prototype

The Modified National Institute of Standards and Technology (MNIST)[5] database is a collection of images of handwritten numbers from 0-9. It contains 60,000 training images and 10,000 testing images. A more recent version includes upper- and lower-case letters. Still, in either case, these images are used to determine how accurately various machine learning models can predict the content of input data. Depending on the model, each image is deconstructed into smaller components that isolate individual patterns that form specific characters. As described earlier, these patterns constitute a 'motif' or 'little phrase' from which a more extensive understanding of a character is established. When trained on enough images, a machine learning model can predict which number or character it is 'seeing' when presented with previously unseen test images. The repeating motifs that define a handwritten number can be aggregated to establish the look of a particular character forming a leitmotif or theme representing a person's handwriting style. This structure or mechanism is prevalent throughout the field and informs the development of the text-to-image vision model(s) discussed below.

Vernacular

In early 2021 Ryan Murdock released a Colab 'notebook' called Big Sleep that combined OpenAI's recently published Contrastive Language-Image Pre-training (CLIP)[6] with Google DeepMind's BigGAN.[7] Around the same time, Katherine Crowson released a similar notebook that combined VQGAN[8] with CLIP. These notebooks and the many variations they inspired are paradigmatic examples of the power of theme augmenting mood into a vernacular, creative and imaginative discourse.

Before these notebooks, the most popular techniques worked with image models that associated pictures with labels; for example, an image of an apple was labeled 'apple, red', or a picture of a desk was labelled 'desk, office'. Vision models were used to identify features in an environment, cars on the road, people in a crowd, parts on an assembly line, and more. In 2016 a group of researchers from the University

---

[5] Y. Lecun, L. Bottou, Y. Bengio, & P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE* 86, 11 (1998): 2278–2324.

[6] Alec Radford, et al. "Learning Transferable Visual Models from Natural Language Supervision," *ArXiv* 2103.00020 [Cs], 2021, http://arxiv.org/abs/2103.00020, acc. on March 6, 2023.

[7] Andrew Brock, Jeff Donahue, and Karen Simonyan, "Large Scale GAN Training for High Fidelity Natural Image Synthesis," *ArXiv* 1809.11096 [Cs, Stat], 2019, http://arxiv.org/abs/1809.11096, acc. on March 6, 2023.

[8] Patrick Esser, Robin Rombach, and Björn Ommer, "Taming Transformers for High-Resolution Image Synthesis," *ArXiv*, 2012.09841 [Cs], 2021, http://arxiv.org/abs/2012.09841, acc. on March 6, 2023.

of Toronto published the image model, alignDRAW[9] trained on Microsoft's COCO[10] dataset (as well as MNIST) to produce images from natural language descriptions. Murdock and Crowson's motivations were similar, and they recognized that the CLIP image/caption model could be used to guide the image distillation process of BigGAN and VQGAN, respectively. Contemporary artistic techniques such as StyleGAN[11] allowed one to produce images that mimicked the style of a famous artist or mirrored the features of a model trained on a specific dataset. Still, it was difficult to steer the output towards something an artist might imagine. Natural language input was exciting to artists because now, the concept and mood of work could be manipulated through the creative construction of a text prompt. By late 2021 there were multiple derivations of the CLIP-guided or text-to-image process that incorporated alternative image generation techniques. Open-source GitHub repositories with associated Google Colab notebooks stimulated a cottage industry of artists producing images and publishing them on social media. 'Prompt engineering' resources and guides to assist stylistic choices were published in blogs and Google Docs (Figure 1). Many techniques spawned their own Reddit channel, where artists posted images and shared their prompts and settings (Figure 2). Immediate feedback and technical support were often available via Discord groups, sometimes populated by the originators of the notebook or repository.

In April of 2022, OpenAI published as a closed beta a revised version of DALL•E called DALL•E 2.[12] Boris Dayma and Pedro Cuenca had previously produced an open-source derivative of DALL•E called DALL•E Mini[13] that went viral in June of 2022, receiving national media coverage. Soon after that, DALL•E 2 was released to the public, for a charge, in July. Similar to DALL•E mini, an upstart created by members of the guided-diffusion community developed a commercial product called Midjourney (2022). The open-source Colab notebooks associated with many non-commercial text-to-image generators require some technical skill. Depending on the desired image resolution and choice of image models, the user must purchase a relatively expensive 'Pro' membership to have enough CPU and GPU memory to function. DALL•E mini and Midjourney simplify this by providing an invisible back-end hardware solution that allows the user to focus on prompt design with minimal configuration.

Many text-to-image works were produced for the now not-so-burgeoning NFT market. Articles on the future of art and art careers when practically anyone

---

[9] Elman Mansimov, Emilio Parisotto, Jimmy Lei Ba, and Ruslan Salakhutdinov, "Generating Images from Captions with Attention," *ArXiv* 1511.02793, 2016, http://arxiv.org/abs/1511.02793, acc. on March 6, 2023.

[10] Tsung-Yi Lin, et al., "Microsoft COCO: Common Objects in Context," *ArXiv* 1405.0312, 2015.

[11] Tero Karras, Samuli Laine, and Timo Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks," *ArXiv* 1812.04948; Version 3, 2019.

[12] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen, "Hierarchical Text-Conditional Image Generation with CLIP Latent," *ArXiv* 2204.06125, 2022. http://arxiv.org/abs/2204.06125, acc. on March 6, 2023.

[13] Bors Dayma, et al., *DALL·E Mini*, 2021.

can create an image without hiring, for example, an album artist or graphic designer, have become popular. Clearly, a vernacular image practice and discourse has emerged that is entirely new and expansive in scope. The public availability of Google's Colab notebook system and cloud GPUs was critical to this development. Recently, it appeared Google and OpenAI are beginning to restrict public access in ways that could undermine vernacular engagement with this technology. In addition to access, they are implementing content filters that restrict words that can be used in a prompt. Well known artist David O'Reilly posted his opinion that DALL•E 2 is a scam on Instagram (Figure 3). For now, a burgeoning community is inventing new images and techniques using AI vision models. A developing discourse asks essential questions about this technology, who it serves, and what it means. This conversation could only emerge under the conditions made possible by fundamental changes in the apparatus.

Mood

Heidegger and Lacan identified anxiety as the primary mood of modernity. From film to theatre to music to therapy, we deal with this anxiety through fantasy, sublimating or reconciling the attraction/repulsion of uncertainty. The history of science fiction is a testament to the power of fantasy in our attempt to imagine the unknown. Contemporary AI image practice shares this tendency. Indeed, many images produced using these techniques are what one would categorize as fantasy images. For example, a simple Google image search for 'disco diffusion fantasy' returns thousands (Figure 4). These images are often inspired by a traditional fantasy artist used as a direct stylistic reference in the prompt, through the selective use of keywords that tend to produce fantasy imagery or simply via the concepts embodied in the text prompt itself. The *Weird Wonderful AI Art* blog provides Disco Diffusion 70+ Artist Studies[14] which, as of July 2022, highlights 662 artists. A simple and consistent set of prompts are used for each artist, making it easy to see the aesthetic effect (mood) contributed to the diffusion process. The site also includes sections devoted to Disco Diffusion Modifiers[15] and Anything Punk Modifiers for AI Art,[16] each providing keywords and sample image representations of their associated stylistic effects on the final output.

The success of AI image practice is not surprising. There is a vacuum of creativity in our daily lives. People who never considered themselves artists, typically because of a lack of technical training and the common belief that art requires technical mastery, can now enjoy, and seek the optimal mood, flow.

---

[14] Harmeet, *Disco Diffusion 70+ Artist Studies | Weird Wonderful AI Art*, February 26, 2022, https://weirdwonderfulai.art/resources/disco-diffusion-70-plus-artist-studies/, acc. on March 6, 2023.

[15] Harmeet, *Disco Diffusion Modifiers | Weird Wonderful AI Art*, March 25, 2022, https://weirdwonderfulai.art/resources/disco-diffusion-modifiers/, acc. on March 6, 2023.

[16] Harmeet, *Anything Punk Modifiers for AI Art | Weird Wonderful AI Art*, March 7, 2022, https://weirdwonderfulai.art/resources/anything-punk-modifiers-for-ai-art/, acc. on March 6, 2023.

Case Study – *Dissipative Off-ramps*

Given the wide range of text-to-image work and the fact that much of it is posted somewhat anonymously online, it is difficult to give proper attribution to artists. Similarly, it can be difficult to access their detailed working processes to reveal the mechanisms of theme at work in the output. In the sections below, we will discuss a case study called *Dissipative Off-ramps*. The artwork was conceived as an AI-generated traversal of the United States in the spirit of the classic 'road trip' from the East coast to the West. The goal of *Dissipative Off-ramps* was to evoke a passion for exploration constantly subverted by entropy or disaster; as much as we want to move forward, we keep repeating the same mistakes. In that sense, the emotion a sense of the sublime, or blues. It consists of ten 'scenes', each of which is guided by individual text prompts, which we will discuss concerning the topics of 'theme' and 'mood'.

Disco Diffusion

Dissipative Off-ramps is a two-minute, five-second animation produced entirely using Disco Diffusion v5.4[17] Disco Diffusion bills itself as "A Frankensteinian amalgamation of notebooks, models, and techniques for the generation of AI Art and Animations." Like many others, it is based on Katherine Crowson's original notebook. It uses her fine-tuned 512x512 diffusion model produced with OpenAI's guided diffusion research documented in Diffusion Models Beat GANs on Image Synthesis[18] The Disco Diffusion source code can be freely accessed via GitHub and is supported by a large community on r/DiscoDiffusion and Discord. The Reddit forum was created on February 9, 2022, and as of July 21, 2022, it has 8,800 members. In addition, three to four thousand simultaneous users regularly occupy the Discord server.

Most of those working with Disco Diffusion and similar systems will launch or download the Colab notebook from the repository and save it to their Google drive. In that scenario, the software depends on access to either the free or paid versions of Google Colab for the necessary access to GPU resources. Disco Diffusion can also be run 'locally', for example, if one owns a computer with an Nvidia GPU or has other means of accessing GPUs. For *Dissipative Off-ramps*, we had access to the University of Florida's HiPerGator AI DGX A100 SuperPod (*HiPerGator – Research Computing – University of Florida*, n.d.)[19]. Each A100 has 80GB of memory, allowing us to experiment with all the available models, even simultaneously, that are made accessible in the Disco Diffusion notebook. For *Dissipative Off-ramps*, we typically consumed approximately 40 GB of GPU memory. While it takes more configuration to run

---

[17] K. Crowson, et al., *Disco Diffusion* [Jupyter Notebook], alembics, https://github.com/alembics/disco-diffusion (Original work published 2022), acc. on March 6, 2023.

[18] Prafulla Dhariwal and Alex Nichol, "Diffusion Models Beat GANs on Image Synthesis," *ArXiv* 2105.05233 [Cs, Stat], 2021, http://arxiv.org/abs/2105.05233, acc. on March 6, 2023.

[19] *HiPerGator – Research Computing – University of Florida.* (n.d.), https://www.rc.ufl.edu/about/hipergator/, acc. on July 24, 2022.

locally, the initial setup time is offset by dependable access to plenty of GPU memory and fewer limits on the amount of time needed to produce multi-frame animation. At a resolution of 576 x 1024 pixels, each frame took approximately 12 minutes to render over 300 hours or 12.5 days to reach the final runtime of 1499 frames. Using a 2013 Mac Pro with dual AMD graphics for approximately 24 hours, UHD resolution frames were produced using an AI image restoration tool called Real-ESRGAN.[20] Of course, that does not factor in the countless hours selecting models and tuning parameters to achieve the intended aesthetic and rendering features before commencing with the final work.

General configuration

Disco Diffusion is complex in that it exposes, depending on the version, more than 80 configuration parameters. Access to this complexity is a crucial advantage of using an open-source system rather than the closed, typically commercial products that have arisen recently. For this paper, discussing each parameter in detail is not necessary. However, online guides, often with cause-and-effect graphics, are indispensable in this regard. Our primary concern is the configuration parameters that direct or affect 'theme' and 'mood'. Many of these parameters interact with one another, and a process of trial and error is necessary to converge on the ideal setup for a particular work. The combinations chosen to determine the aesthetic the artist is attempting to produce. While the term 'prompt engineering' has become popular to describe the strategic use of keywords and construction of textual hints for the AI, the skillful manipulation of the actual engineering defined by these configuration parameters is equally important to set a mood for the work. Many traditional image composition techniques are available, such as color range, saturation, field of view, and more.

One of the more important general configuration choices is the selection of image models. This is a key point at which 'theme' plays a role. Disco Diffusion allows one to load custom vision models or select from several popular choices. For this project we chose to enable ViTB32, ViTB16, ViTL14_336px, RN50, and RN50x64. The selection and classification of images with which to construct the final image is determined by the choice of vision model(s). Models with prefixes 'ViT' are Vision Transformer models that have recently (2022) been shown to be up to four times as accurate and efficient as prior Convolutional Neural Network (CNN) models.[21] These models are pre-trained on the ImageNet-21k dataset consisting of over 14 million images organized into 21,000 categories.[22] In this case, we are loading three vision transformers of various resolutions and patch sizes into our system. Because ViTL14_336px is composed of relatively high-resolution images (336 pixels is considered large), it

---

[20] X. Wang et al., "Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data," *International Conference on Computer Vision Workshops (ICCVW)*, 2021.

[21] Alexey Dosovitskiy, et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," *ArXiv* 2010.11929, 2021.

[22] Tal Ridnik, et al., "ImageNet-21K Pretraining for the Masses," *ArXiv* 2104.10972, 2021.

can consume more than 20 GB of video memory. RN50 and RN50x64 are ResNet models,[23] the more traditional computer vision CNN models. We experimented with several combinations of models and ultimately settled on these as demonstrating the greatest potential to recognize the concepts and thematic or artistic influences necessary in our scenes.

Another critical parameter affecting a particular session's theming capability is 'clip_guidance_scale'. This parameter sets a relationship between the CLIP interpretation of the text prompt and how quickly each diffusion step should move towards the prompt. The scale depends on the image resolution, so it takes some testing to get this right. Otherwise, output images will be distorted if too high or not reflect the prompt very well if too low. We found that 75000 worked well given this particular set of conditions.

Dissipative Off-ramps utilized Disco Diffusion's 'Turbo Mode', a feature that allows one to skip diffusion steps to save time selectively. Turbo is designed to assist those doing 3D animation, recognizing that the temporal aspects of a 12 or 24-frame-per-second motion image do not require every frame to undergo the complete diffusion process. We set the number of steps to 500 and allowed it to skip up to 250. The diffusion process produces images by working from noise toward the final output. Part of that process is to break the image into smaller patches that can be controlled via several scheduling parameters. These parameters are essential to create sharpness and detail and allow the latent space to 'emerge' during the process of imaging. With 1000 diffusion steps, for the overview pass, the image was broken into 24 patches for the first 400 passes and 4 for the final 600. For the innercuts, the first 400 passes were cut four times, with the last 600 seeing 24 innercuts.

Prompt engineering

The construction of a text prompt is vital in the text-to-image process. The goal is to be as descriptive as possible, imagining what a computer might need to make sense of our desires. In this case, after a period of trial and error, we converged on a set of prompts that invoked the mood intended for the work. The prompts are designed to be evocative, not literal, and are structured in a JSON format with commas separating sections that are weighted in importance. Disco Diffusion allows one to have multiple prompts prefixed by a frame number. The engine will apply the prompt to the entire range of frames within a block, then transition to the next block at the designated frame. For *Dissipative Off-ramps*, each of the ten prompts, or scenes, was structured similarly:

Segment 1 – describes the primary content of the scene.
Segment 2 – describes the physical environment.
Segment 3 – identifies a preferred color palette.
Segment 4 – selects an artist or stylistic referent.

---

[23] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep Residual Learning for Image Recognition," *ArXiv* 1512.03385, 2015.

Segment 5 – keywords known to stimulate an imaging technique.

Segment 6 – depth of focus control.

There is nothing that requires this format. We arrived at this method after a period of trial and error. Each of these segments is biased by adding a weighting factor at the end of each line. The weighting factors communicate to the engine how important a particular segment is in the context of the overall scene. Below, the reader can view a sample still frame and the associated prompt used to generate the scene. The work is intended to be viewed on a UHD monitor oriented vertically. One can view the animation formatted for normal horizontal viewing on YouTube:[24]

Scene 1 [Figure 5]

"0": [

"A photograph of a crowded sea port with sailboats and freighters and seagulls flying amidst billowing clouds set against a beautiful orange sunrise:10",

"swampy environment with moss covered trees and dusty roads as we drive by:8",

"orange color scheme:5",

"by Gustave Le Gray:8",

"very intricate, very detailed:7",

"dof, depth of field:-10"

],

Scene 2 [Figure 6]

"150": [

"A close-up photograph of the space shuttle Challenger launching from the shore and streaking up into a roiling billowing sky as it explodes, debris falls off the rocket and crashing to the water:10",

"orange and blue color scheme:6",

"by NASA:8",

"8K, hyperrealism, Unreal Engine, very intricate, very detailed:8",

"dof, depth of field:-10"

],

Scene 3 [Figure 7]

"300": [

"A photograph of beautiful Appalachian Blue Ridge forest with summer cottages, bubbling streams, Appalachian Trail wildflowers and clear blue sky with jet contrails:10",

"mixed forest with rolling hills, wildflowers, streams and asphalt roads as we drive by:8",

"yellow and blue green color scheme:5",

---

[24] Jack Stenner (Director), *Dissipative Off-ramps (2022) by Jack Stenner*, https://www.youtube.com/watch?v=t3HTwt3uInQ, July 26, 2022.

"by National Geographic, by Robert Wojtowicz :8",
"8K, hyperrealism, Unreal Engine, very intricate, very detailed:7",
"dof, depth of field:-10"
    ],
Scene 4 [Figure 8]
    "450": [
        "A photograph of a West Virginia coal mine in a town like Nuttallburg or Kaymoor with huge cranes, draglines and mine shafts and decaying shanties with abandoned rusty metal storage buildings under an ominous storm cloud:10",
        "coal mines, mountaintop removal, strip mines, heroin needles, desolate black coal and dirty roads:9",
        "dark rusty red and orange color scheme:7",
        "by Bernd and Hilla Becher:9",
        "8K, hyperrealism, Unreal Engine, very intricate, very detailed:7",
        "dof, depth of field:-10"
    ],
Scene 5 [Figure 9]
    "600": [
        "A beautiful photograph of sprawling Kansas farmland with fields of sunflowers, farm houses and barns and red tractors with massive round bales of hay as far as the eye can see with billowing blue and white clouds:10",
        "sunflower fields photograph, farmland, sunflowers and dirt roads as we drive by:8",
        "blue and yellow color scheme:5",
        "photograph, sunflower field:8",
        "high resolution, very intricate, very detailed:7",
        "dof, depth of field:-10"
    ],
Scene 6 [Figure 10]
    "750": [
        "A photograph of a massive thunderstorm storm with huge grain silos breaking open and grain spilling everywhere. A drainage culvert flows toxic sludge as farmland becomes barren and bleak while a tornado extends from the rain clouds and destroys trailer houses:10",
        "grain silos and conveyor belts with dusty dirt roads and broken barbed wire fences as we drive by:8",
        "brown color scheme:5",
        "by Dorothea Lange:8",
        "8K, hyperrealism, Unreal Engine, very intricate, very detailed:7",
        "dof, depth of field:-10"
    ],

Scene 7 [Figure 11]

    "900": [

        "A beautiful photograph of downtown Chicago skyscrapers a bright blue sky and with building reflections shimmering off the surface of a huge lake. Red, blue and yellow balloons blow in the wind and float into the atmosphere:10",

        "yellow color scheme:5",

        "by Walker Evans architectural:8",

        "8K, hyperrealism, Unreal Engine, very intricate, very detailed:7",

        "dof, depth of field:-10"

    ],

Scene 8 [Figure 12]

    "1050": [

        "A photograph of a post-apocalyptic Chernobyl with abandoned streets, a nuclear silo on fire with missiles dropping acid green rain and black clouds in the sky:10",

        "yellow green color scheme:5",

        "by Hisaharu Motoda:8",

        "very intricate, very detailed:7",

        "dof, depth of field:-10"

    ],

Scene 9 [Figure 13]

    "1200": [

        "A photograph of suburban neighborhood with automobiles and mail boxes and trailer parks with abandoned gas stations and roiling black sky:10",

        "abandoned damaged automobiles and dead trees on barren mountains beside asphalt roads as we drive by:8",

        "sepia color scheme:8",

        "by Todd Hido:8",

        "8K, hyperrealism, very intricate, very detailed:7",

        "dof, depth of field:-10"

    ],

Scene 10 [Figure 14]

    "1350": [

        "A detailed photograph of a single 1960 Volkswagen hippie van decaying beside the road in a desert filled with cactus and bomb craters with rockets and an ominous storm cloud in a bright blue sky:10",

        "An abandoned gas station with a broken neon sign:4",

        "Central Road, Apple Valley, CA, from the series Mojave by Markus Altmann:8",

        "very intricate, very detailed:7",

        "dof, depth of field:-10"

    ]

  },

Conclusion

To the uninitiated, it may look as if AI art production is simply a matter of feeding a bot a sentence, and it draws a picture. That may be true concerning several of the black-box commercial products that have been recently released and caught the attention of the press. However, artists working with advanced scientific research published through open-source forums like GitHub and Colab are developing new ways to work with AI in a natively electrate fashion, privileging the affective, visceral aspects of meaning-making. The scene prompts and working processes above demonstrate how artists interface with theme and suggest mood when constructing these works. Motif, leitmotif/theme, and mood are operationalized at multiple levels in the process, and artists are developing methods to direct these mechanisms to realize their intuitions. Motif is embodied in the structure of patterns defining the latent space represented in computer vision models. Theme emerges via techniques to reorganize or extract these patterns from massive datasets that can be fine-tuned for more refined thematic content when needed. The ambition is that theming, applying the poetics of motif, will do for visceral sensation in electracy what conceptual categories did for rational logic in literacy. Theming in digital media augments mood (ambiance) into a power of imagination that is as compelling in metaverse entertainment as scientific proof is in natural law.

References

Brock, Andrew, Jeff Donahue, and Karen Simonyan. "Large Scale GAN Training for High Fidelity Natural Image Synthesis." *ArXiv* 1809.11096 [Cs, Stat], 2019. http://arxiv.org/abs/1809.11096. Accessed on March 6, 2023.

Crawford, Kate. *Atlas of Ai: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale: Yale University Press, 2021.

Crowson, K., D. Russell, Dango233, M. Shepherd, C. Allen, C. Nri, Somnai, A. Letts, & Gandamu. *Disco Diffusion* [Jupyter Notebook]. alembics. https://github.com/alembics/disco-diffusion (Original work published 2022). Accessed on March 6, 2023.

Dayma, Boris, Suraj Patil, P. Cuenca, Khalid Saifullah, Tanishq Abraham, Phúc Lê Khắc, L. Melas, and Ritobrata Ghosh. *DALL·E Mini*, 2021. https://doi.org/10.5281/zenodo.5146400

Dhariwal, Prafulla and Alex Nichol. "Diffusion Models Beat GANs on Image Synthesis." *ArXiv* 2105.05233 [Cs, Stat], 2021. http://arxiv.org/abs/2105.05233. Accessed on March 6, 2023.

Dosovitskiy, Alexey, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale." *ArXiv* 2010.11929, 2021. https://doi.org/10.48550/arXiv.2010.11929

Esser, Patrick, Robin Rombach, and Björn Ommer. "Taming Transformers for High-Resolution Image Synthesis." *ArXiv*, 2012.09841 [Cs], 2021. http://arxiv.org/abs/2012.09841. Accessed on March 6, 2023.

Harmeet. *Disco Diffusion 70+ Artist Studies | Weird Wonderful AI Art*. https://weirdwonderfulai.art/resources/disco-diffusion-70-plus-artist-studies/, February 26, 2022. Accessed on March 6, 2023.

Harmeet. *Anything Punk Modifiers for AI Art | Weird Wonderful AI Art*. https://weirdwonderfulai.art/resources/anything-punk-modifiers-for-ai-art/, March 7, 2022. Accessed on March 6, 2023.

Harmeet. *Disco Diffusion Modifiers | Weird Wonderful AI Art*. https://weirdwonderfulai.art/resources/disco-diffusion-modifiers/. March 25, 2022. Accessed on March 6, 2023.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep Residual Learning for Image Recognition." *ArXiv* 1512.03385, 2015. https://doi.org/10.48550/arXiv.1512.03385

HiPerGator – Research Computing – University of Florida. (n.d.). https://www.rc.ufl.edu/about/hipergator/. Accessed on July 24, 2022.

Jack Stenner (Director). *Dissipative Off-ramps (2022) by Jack Stenner*. https://www.youtube.com/watch?v=t3HTwt3uInQ. July 26, 2022.

Kantayya, S. (Director). *Coded Bias*, 2020. https://www.codedbias.com. Accessed on March 6, 2023.

Karras, Tero, Samuli Laine, and Timo Aila. "A Style-Based Generator Architecture for Generative Adversarial Networks." *ArXiv* 1812.04948; Version 3, 2019. https://doi.org/10.48550/arXiv.1812.04948

Lecun, Y., L. Bottou, Y. Bengio, & P. Haffner. "Gradient-based learning applied to document recognition." *Proceedings of the IEEE* 86, 11 (1998): 2278–2324. https://doi.org/10.1109/5.726791

Lin, Tsung-Yi, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. "Microsoft COCO: Common Objects in Context." *ArXiv* 1405.0312, 2015. https://doi.org/10.48550/arXiv.1405.0312

Mansimov, Elman, Emilio Parisotto, Jimmy Lei Ba, and Ruslan Salakhutdinov. "Generating Images from Captions with Attention." *ArXiv* 1511.02793, 2016. http://arxiv.org/abs/1511.02793. Accessed on March 6, 2023.

Mordvintsev, Alexander. "Inceptionism: Going Deeper into Neural Networks." *Google AI Blog*, June 17, 2015. http://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html. Accessed on March 6, 2023.

Radford, Alec et al. "Learning Transferable Visual Models from Natural Language Supervision." *ArXiv* 2103.00020 [Cs], 2021. http://arxiv.org/abs/2103.00020. Accessed on March 6, 2023.

Ramesh, Aditya, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. "Hierarchical Text-Conditional Image Generation with CLIP Latents." *ArXiv* 2204.06125, 2022. http://arxiv.org/abs/2204.06125. Accessed on March 6, 2023.

Ridnik, Tal, Emanuel Ben-Baruch, Asaf Noy, and Lihi Zelnik-Manor. "ImageNet-21K Pretraining for the Masses." *ArXiv* 2104.10972, 2021. https://doi.org/10.48550/arXiv.2104.10972

Wang, X., Xie, L., Dong, C., & Shan, Y. "Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data." *International Conference on Computer Vision Workshops (ICCVW)*, 2021.

Zuboff, Shoshana. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (First Trade Paperback Edition). PublicAffairs, 2020.
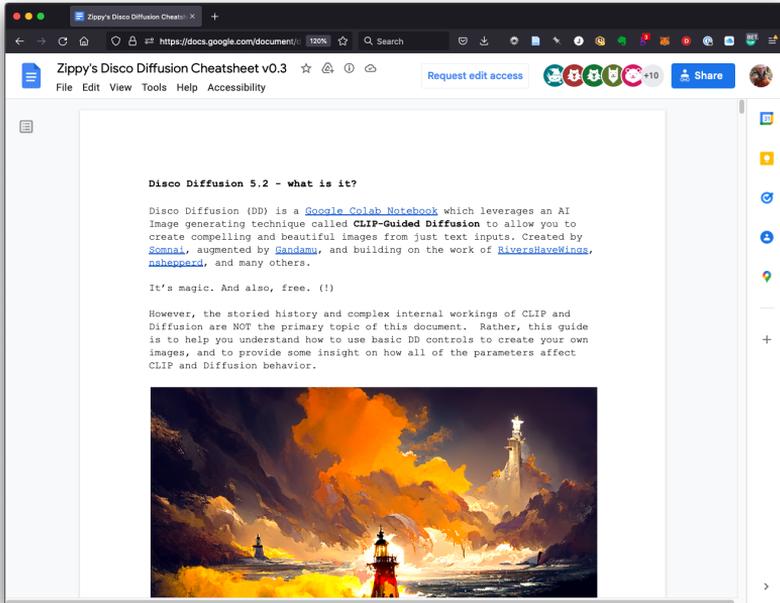
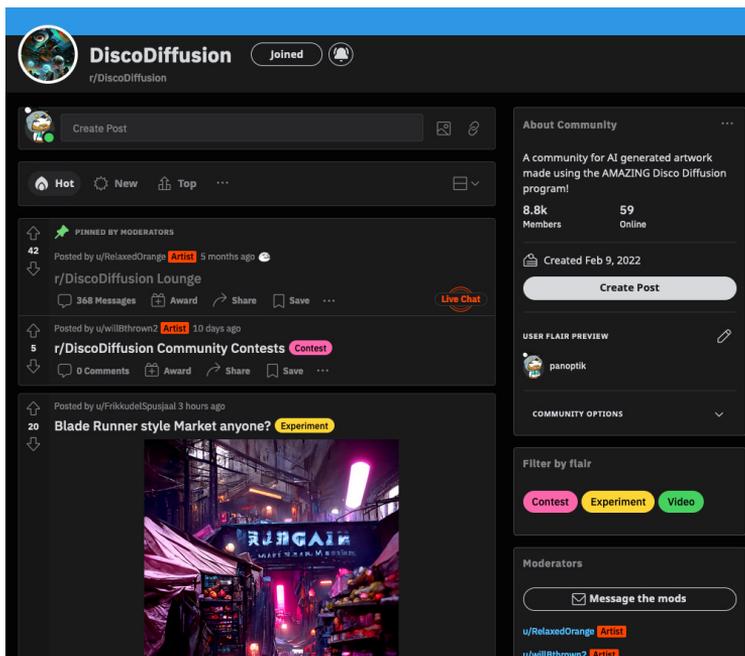**Figure 1**. Zippy's Disco Diffusion Cheatsheet Shared Google Doc

Figure 2. r/DiscoDiffusion Reddit group with 8,800 members.
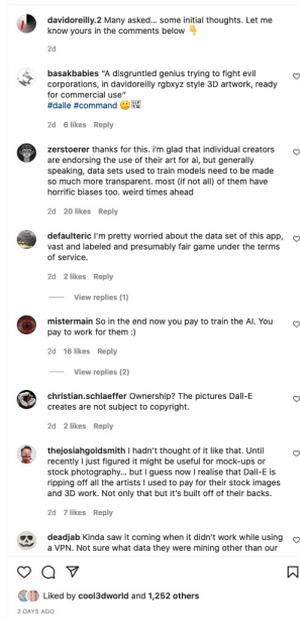
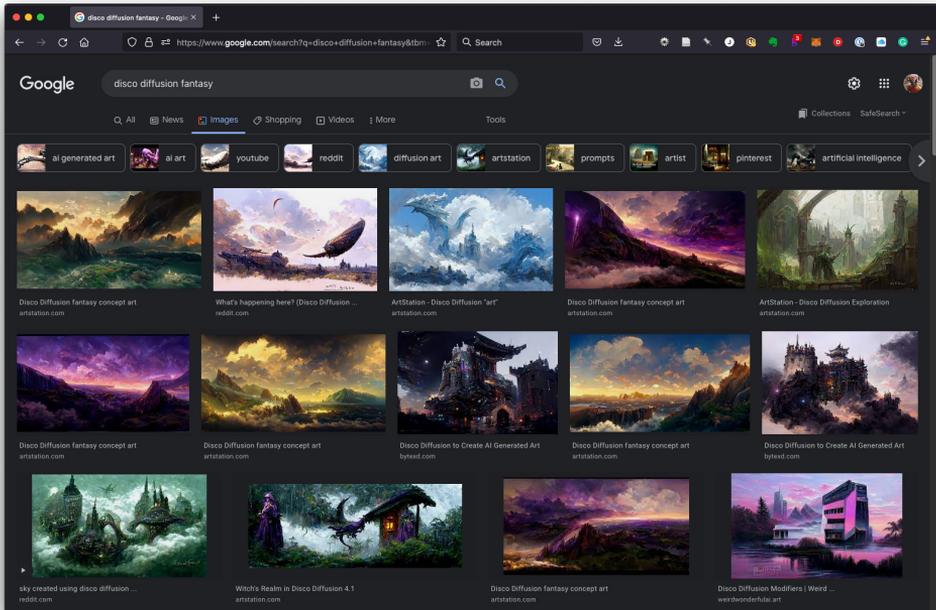**Figure 3**. Artist David O'Reilly on DALL•E, Instagram, July 22, 2022.

Figure 4. Google Images search for, "Disco Diffusion fantasy".

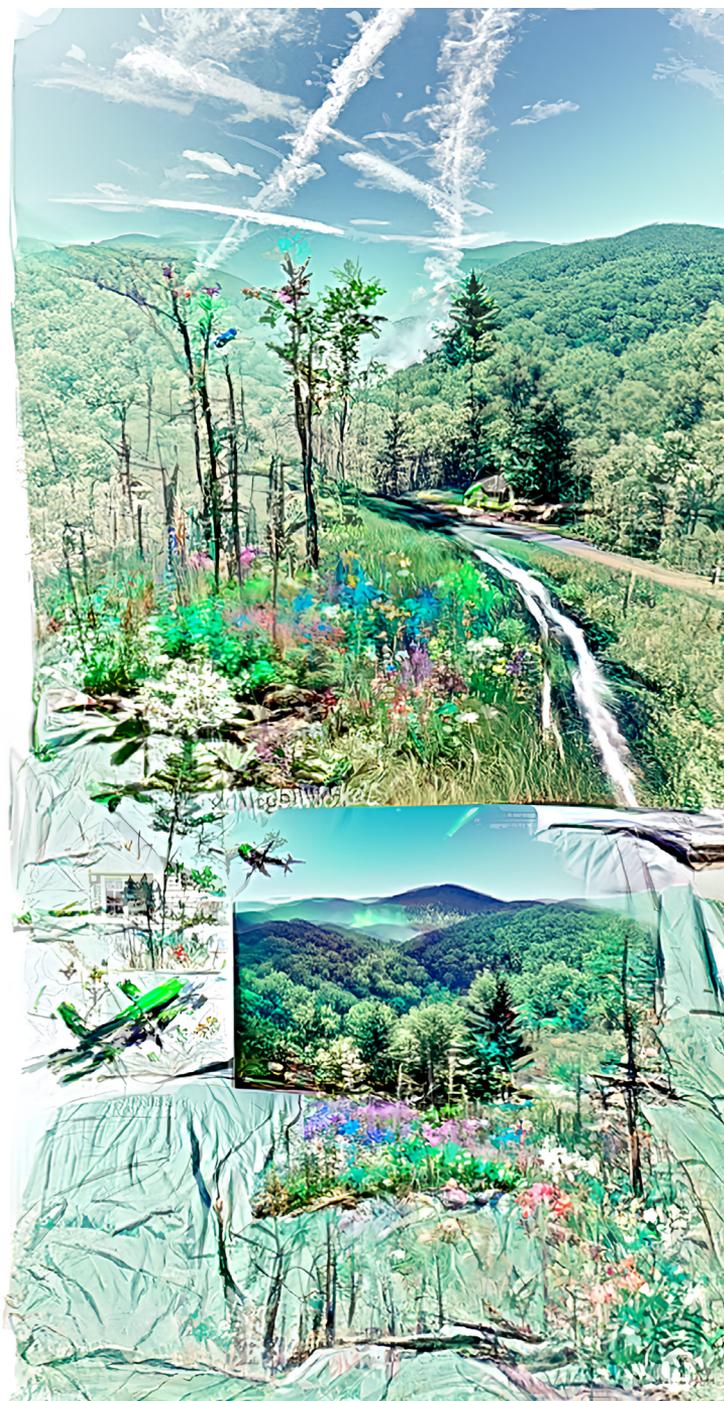**Figure 5**. Scene 1, sample frame.

Figure 6. Scene 2, sample frame.

Figure 7. Scene 3, sample frame.

Figure 8. Scene 4, sample frame.

Figure 9. Scene 5, sample frame.

Figure 10. Scene 6, sample frame.

Figure 11. Scene 7, sample frame.

Figure 12. Scene 8, sample frame.

Figure 13. Scene 9, sample frame.

Figure 14. Scene 10, sample frame.

Supplemental material

Dissipative Off-ramps, YouTube video: https://www.youtube.com/watch?v=t3H-Twt3uInQ